STATISTICS
WITH THE
GRAPHING CALCULATOR TI83Plus

Guido Herweyers
T3 - Europe
Belgium
email: guido.herweyers@khbo.be

1) **Runs in binary random numbers**

2) **Approximate the probability distribution and the mean of a random variable by simulation**

3) **Simulate a game of chance**

4) **Investigate a population**

5) **Draw samples from a population**

6) **Normal approximation of a binomial distribution**

7) **Confidence Intervals**

8) **F-test for comparing two standard deviations**

9) **Modeling data**

1) Runs in binary random numbers

Toss a coin 10 times. What is the probability of a longest "run" of 5 consecutive heads or tails?

We will estimate this probability by simulating many times the experiment, consisting of tossing a coin 10 times, with the TI-83; we therefore represent head by 1 and tail by 0.

First we simulate the experiment 5 times by generating a list of 50 random bits with randInt(Ø,1,5Ø). The result is, in groups of 10 bits:

0101000001 1110110011 0111100001 0011000100 1101011001.

The first group is the only group with a longest run of length 5. A rough estimation for the probability of a longest run of 5 consecutive heads or tails is given by the relative frequency 1/5 = 20 %.
We write a program for simulating the experiment 50 times, the basic idea is:

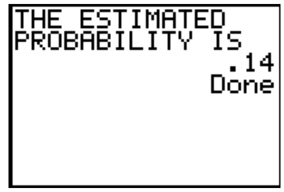
$0 \rightarrow T$ (T counts the number of groups of 10 bits with a longest run of length 5)

Repeat 50 times :
Generate a sequence of 10 bits in L1
Determine the length S of the longest run in L1
Increase T with 1 if S = 5

Calculate the relative frequency T/50

```
PROGRAM:RUN
ClrHome                                 clear the home screen
Ø→T                                     initialisation of T
For(I,1,5Ø)                             repeat 50 times
 randInt(Ø,1,1Ø)→L1                      10 bits in L1
 1→S                                     initialisation of S
 1→K                                     K counts the length of each run in L1
 1→J                                     J is the position of the bit in L1
 While J<1Ø
  While (L1(J)= L1(min({J+1,1Ø})) and J<1Ø)
   K+1→K                                  increase K with 1
   J+1→J                                  increase J with 1 within a run
  End                                     end of second While
  J+1→J                                   increase J with 1 for a new run
  If K>S
  K→S                                     S remembers the longest run
  1→K                                     put K = 1 for the next run
 End                                      end of first While
 If S=5                                   increase T with 1 if S=5
 T+1→T
End                                       end For-loop
Disp "THE ESTIMATED"
Disp "PROBABILITY IS"
Disp T/5Ø
```

The result of running the program once gives 0.14.
The exact probability is $\frac{63}{512} = 0.123$.



A longest run of 5 consecutive heads or tails in tossing a coin 10 times is not that seldom!

<u>Class task</u>

Estimate the probability by combining the results of the students in your class.

Change the program such that it asks at the start how many times you want to repeat the experiment (always 50 times in the above program).

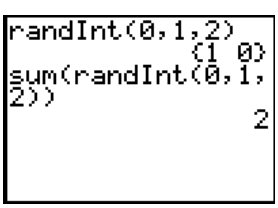
Estimate the probability of a longest run of length 3 (exact : 185/512 = 0.361) and the probability of a longest run of length at least 5 (exact : 111/512 = 0.217).

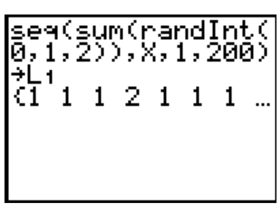
Verify the exact probability distribution of a longest run of heads (1) or tails (0) with length l:

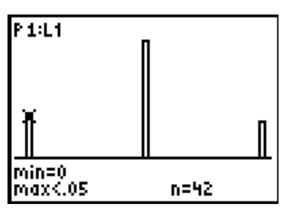
l	1	2	3	4	5	6	7	8	9	10
$P(l)$	$\frac{1}{512}$	$\frac{88}{512}$	$\frac{185}{512}$	$\frac{127}{512}$	$\frac{63}{512}$	$\frac{28}{512}$	$\frac{12}{512}$	$\frac{5}{512}$	$\frac{2}{512}$	$\frac{1}{512}$

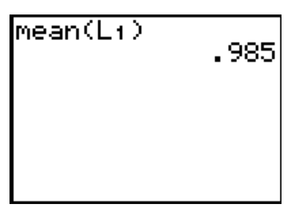
2) Approximate the probability distribution and the mean of a random variable by simulation.

Toss a coin twice and let X = number of heads. What is the probability distribution of X?
Simulate the experiment "toss a coin twice" 200 times, noting down the number of heads in list **L1**. Code head with 1 and tail with 0. Plot the frequency distribution.









The simulation gives 42 times 0, 119 times 1 and 39 times 2 as number of heads. The mean is $0\cdot\frac{42}{200}+1\cdot\frac{119}{200}+2\cdot\frac{39}{200}=0.985$, an estimation of the theoretical value $E(X)=1$.

The relative frequencies 0.21, 0.595, 0.195 approximate the probability distribution 0.25, 0.5, 0.25 of the random variable *X*.

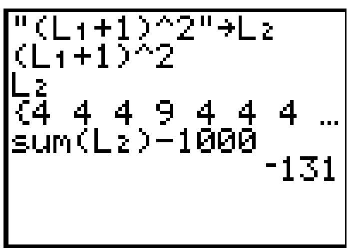
Class task

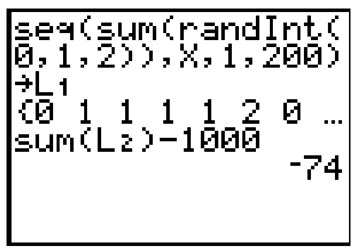
Combine the simulations of all the students. What do you expect of the relative frequency distribution?

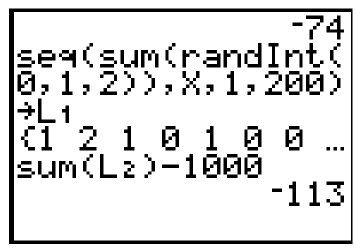
3) Simulate a game of chance

Toss a coin twice and count the number of heads, add 1 and square the result. This is the amount you get in dollars. But you have to pay 5 dollars each game. Are you prepared to play this game a whole evening?

We simulate the game with list `L1` of the preceding simulation. The received amounts appear in the coupled list `L2` = "$(L1 + 1)^2$". Check your loss or profit after 200 games.







You lose 131 dollars with the first simulation. A second and third simulation give a loss of 74 en 113 dollars after 200 games. Don't play this game!

let X = number of heads in tossing a coin twice, the random variable $Y=(X+1)^2-5$ is the profit per game.

Y takes the values $-4,\ -1,\ 4$ with probabilities $\frac{1}{4},\ \frac{1}{2},\ \frac{1}{4}$.

The mean of Y is $E(Y)=-4\cdot\frac{1}{4}+(-1)\cdot\frac{1}{2}+4\cdot\frac{1}{4}=-0.5$.

This is the mean "profit" in the long run.

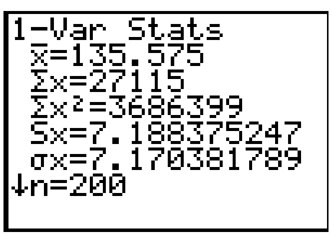
After 200 games, you can expect a loss in the neighbourhood of 100 dollars. If we combine the three simulations, we lose 113 + 131 + 74 = 318 dollars after 600 games. This is close to the theoretical prediction of $600\cdot 0.5=300$.

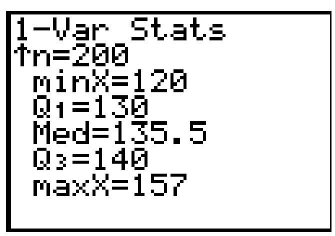
4) Investigate a population

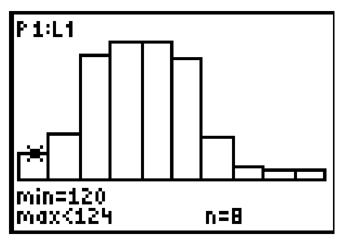
Consider the heights of 200 children (in cm) , age 10, as a population:

120	123	151	142	137	128	133	142	136	137
144	126	135	142	135	134	140	149	137	129
128	140	129	137	141	137	135	129	133	139
132	125	124	132	129	139	132	145	140	138
137	133	137	138	131	137	131	127	134	134
150	140	144	137	133	139	130	141	136	124
130	135	124	122	136	132	133	133	142	127
142	130	135	125	136	132	153	145	131	131
134	145	139	132	136	143	138	141	141	141
136	148	128	137	134	138	130	145	135	141
131	143	146	132	127	129	133	142	157	133
139	128	123	140	140	152	136	125	130	153
130	126	129	157	144	142	128	138	142	135
141	139	132	135	145	134	140	136	138	143
122	141	122	132	136	129	138	130	129	135
134	141	133	128	121	131	137	140	133	135
138	132	140	145	128	140	134	128	146	132
131	142	133	137	126	128	129	124	137	127
139	141	157	146	128	136	130	141	129	143
137	143	139	141	121	131	128	133	136	146

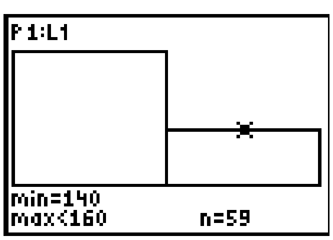
Find the statistics and draw a histogram:



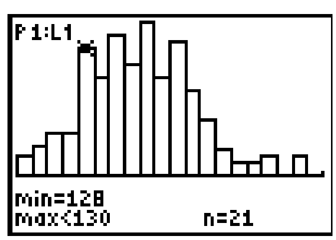




Influence of the class width:

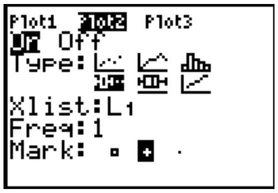


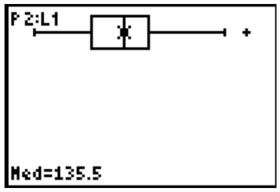
class width 20, too large



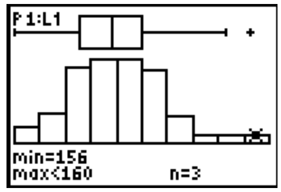
class width 2, too small

boxplot and normal probability plot:

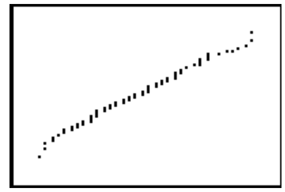




boxplot with outliers



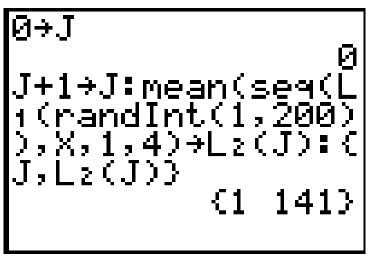
boxplot and histogram

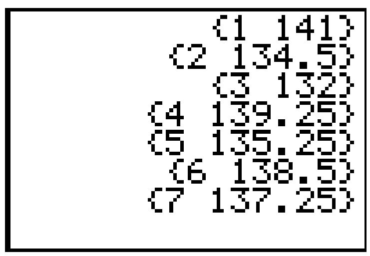

normal probability plot

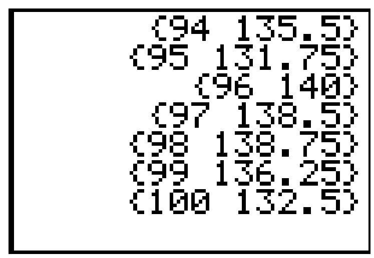
task:
remove the outliers and study the effect on the statistics.

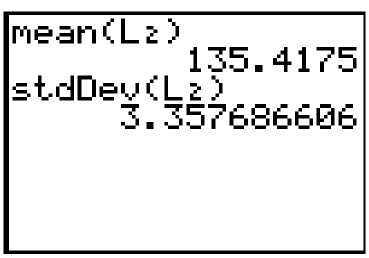
5) Draw samples from a population

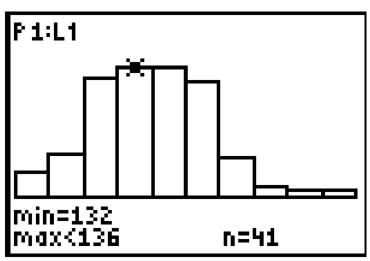
Generate 100 samples of size 4 from the population of 200 heights. Put the means in list L2.
The third sample has mean 132:



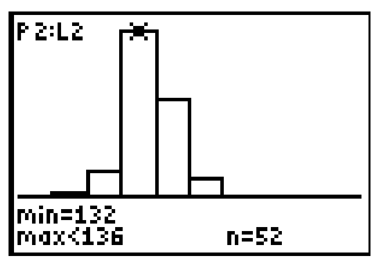








population



sample means

The population has mean $\mu_X = 135.575$.
mean(L2) $= 135.4$ is an estimation of $\mu_{\overline{X}} = \mu_X$

The population has standard deviation $\sigma_X = 7.17$.

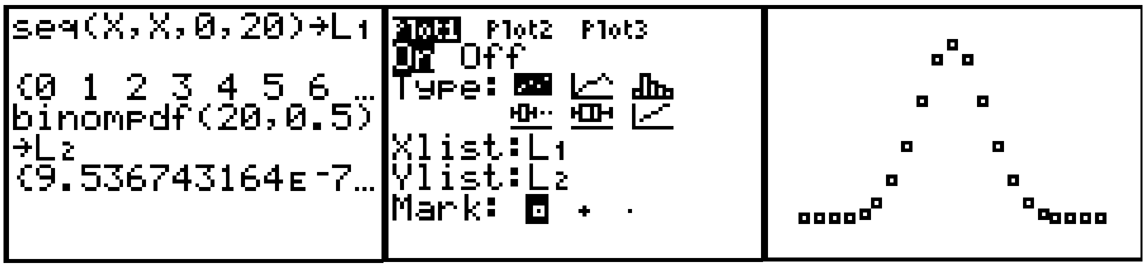
stdDev(L2) $= 3.36$ is an estimation of $\sigma_{\overline{X}} = \dfrac{\sigma_X}{2} = 3.585$

6) Normal approximation of a binomial distribution

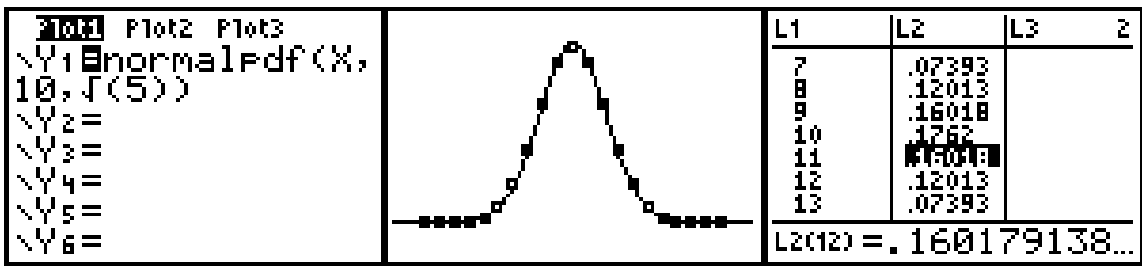
Let X = number of heads in tossing a coin 20 times. X has a binomial distribution with parameters $n = 20$ and $p = 0.5$.

X has mean $\mu = np = 10$ and variance $\sigma^2 = npq = 5$.

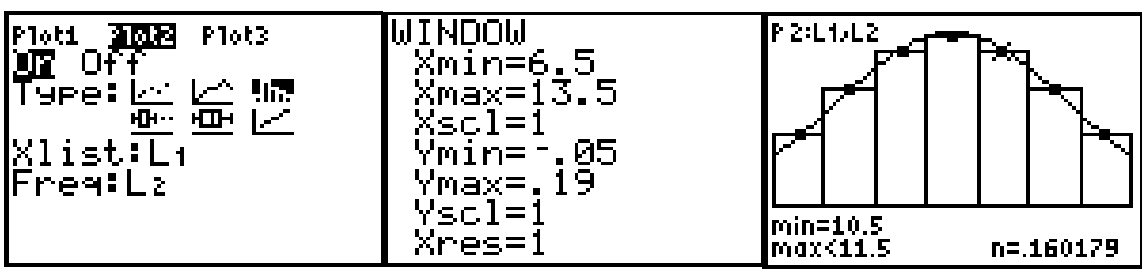
Plot of the probability distribution of X:



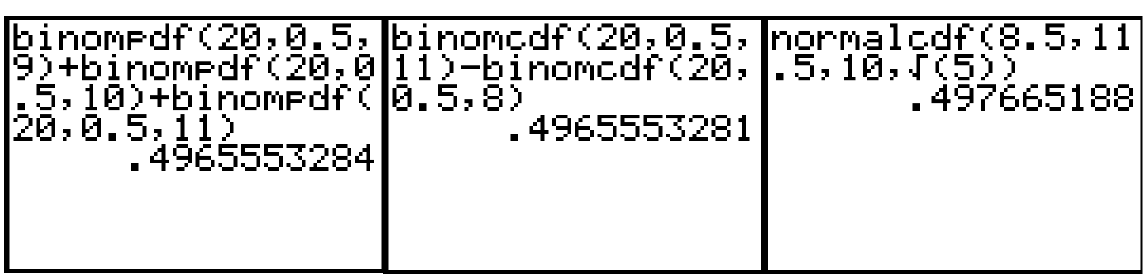
Compare with the normal approximation:



Illustrate the continuity correction:

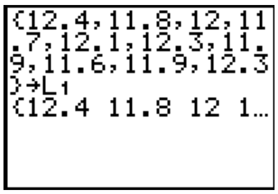


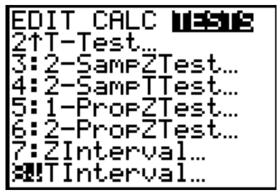
Calculate $P(9 \le X \le 11)$ exact and compare with the normal approximation with continuity correction:

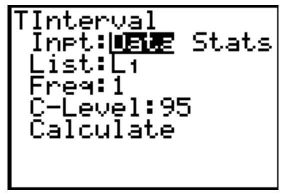


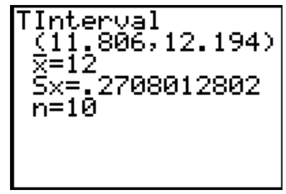
7) Confidence Intervals

Given a sample of size 10 from a normal population with unknown mean μ and standard deviation σ:
12.4   11.8   12.0   11.7   12.1   12.3   11.9   11.6   11.9   12.3
find a 95% confidence interval for the mean of the population:









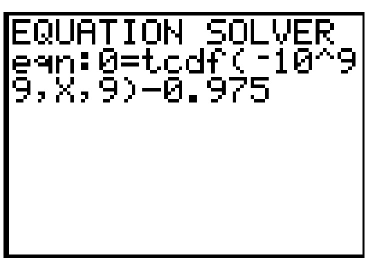
a 95 % confidence interval for μ is [11.806 , 12.194]

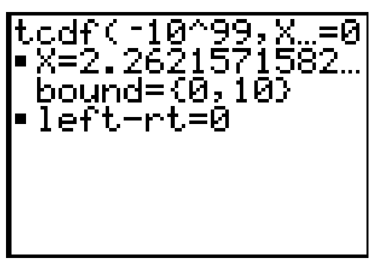
How do we verify this interval?

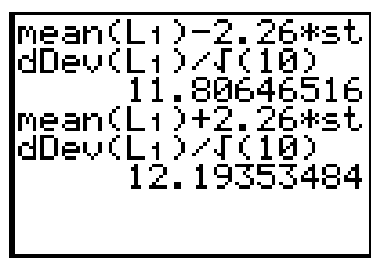
Given $X \sim N(\mu,\sigma)$ and sample mean $\overline{X} = \dfrac{X_1 + X_2 + \ldots X_n}{n}$

then $T = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$ has a t distribution with $(n-1)$ degrees of freedom. Determine t^* such that

$P(-t^* \le T \le t^*) = 0.95$ or $P(\overline{X} - t^* \cdot \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t^* \cdot \frac{S}{\sqrt{n}}) = 0.95$:







$t^* = 2.26$ confidence interval confirmed

8) F-test for comparing two standard deviations

Draw a sample from 2 normal populations, with size $n_1 = 16$ and $n_2 = 9$.
The means and sample standard deviations are:
$\mu_1 = 180$ and $s_1 = 6$, $\mu_2 = 168$ and $s_2 = 4$.

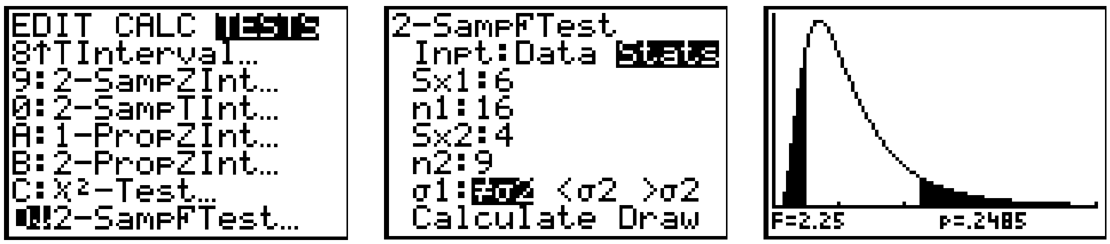
Can you reject the hypothesis of equal population standard deviations at the 5% significance level?

The hypotheses are: $H_0: \ \sigma_1 = \sigma_2$

$H_1: \ \sigma_1 \neq \sigma_2$

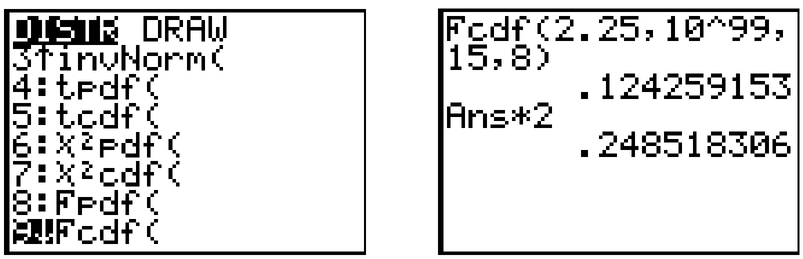
The test statistic $F = \dfrac{S_1^2}{S_2^2}$ has the F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom when H_0 is true.

The p-value = 0.2485 > 0.05 , we cannot reject H_0.



F takes the value 36/16 = 2.25 .

Verify the p-value:

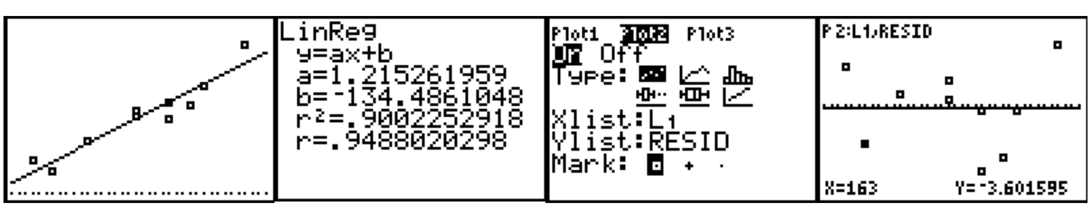


9) Modeling data

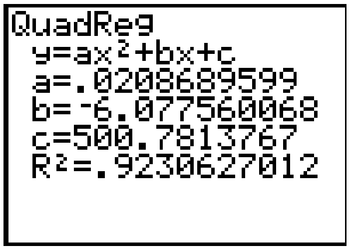
Find an appropiate model for the following data (heights in cm and weights in kg of 10 students).

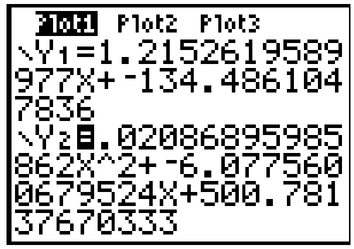
x_i	163	185	180	175	168	175	191	180	160	183
y_i	60	90	78	81	71	79	104	84	64	83

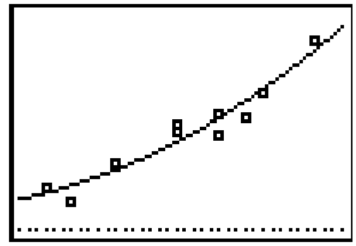
Enter the data in the lists `L1` en `L2`, plot the scatter diagram. A linear model seems to be a good fit. Plot the residuals.



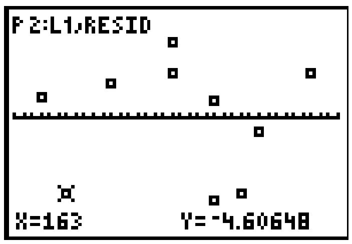
The coefficient of determination $r^2 = 0.90$. This means that 90% of the variation of the observations y_i about the mean $\overline{y}$ is explained by the linear regression model.
Investigate the quadratic model:

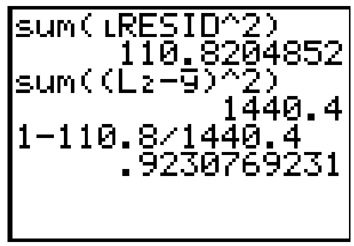






Plot the residuals and verify the value of the determination coefficient $R^2 = 0.923$:





Task:
investigate other models. Verify how, by transformation of the data, linear regression is used to calculate LnReg , ExpReg, PwrReg.

The results for the different models are:

Model	R^2	r^2	r
LinReg(ax+b)		0.900	0.949
QuadReg ($ax^2 + bx + c$)	0.923		
CubicReg ($ax^3 + bx^2 + cx + d$)	0.938		
QuartReg ($ax^4 + bx^3 + cx^2 + dx + e$)	0.963		
LnReg ($a + b.\ln(x)$)		0.892	0.945
ExpReg ($a.b^x$)		0.915	0.957
PwrReg ($a.x^b$)		0.911	0.955

References

1. G. Barret, *Statistics with the TI-83*, Meridian Creative Group, 1997.

2. G. Herweyers, K. Stulens, *Statistiek met een grafisch rekentoestel*, Acco, Leuven, 2000.

3. L. Morgan, *Statistics handbook for the TI-83*, Texas-Instruments, 1997.

4. D.S.Yates, D.S. Moore, G.P. McCabe, *The Practice of Statistics, TI-83 Graphing Calculator Enhanced*, W.H. Freeman and Company, New York, 1999.